

# Data Passing with High Security through Non Infected Nodes in Networks

K. Malarvizhi<sup>1</sup>, Prof. P. Parthasarathy<sup>2</sup>, Dr. S. Shankar<sup>3</sup>

<sup>1</sup>PG Scholar, <sup>2</sup>Assistant Professor, <sup>3</sup>Professor and Head,

<sup>1,2</sup>Akshaya College of Engineering and Technology, Coimbatore, Tamil Nadu, India

<sup>3</sup>Hindusthan College of Engineering and Technology, Coimbatore, Tamil Nadu, India

## ABSTRACT

Path-based network diffusion kernel is its tenability, that it can consider different number of shortest paths in kernel computation. This resembles, vaguely, a Taylor-expansion of network topology to form a diffusion kernel with different orders of expansion. One can extend this key idea to design other network diffusion kernels to approximate other general diffusion models such as SIR the challenge of determining sources is compounded as the true propagation dynamics are typically unknown, and when they have been directly measured, they rarely conform to the assumptions of any of the well-studied models. In this paper introduce a method called Network Infusion (NI) that has been designed to circumvent these issues, making source inference practical for large, complex real world networks. A stateless receiver-based multicast (RBMulticast) protocol that simply uses a list of the multicast members' (e.g., sinks') addresses, embedded in packet headers, to enable receivers to decide the best way to forward the multicast traffic. This protocol, called Receiver-Based Multicast, exploits the knowledge of the geographic locations of the nodes to remove the need for costly state maintenance. The key idea is that to infer the source node in the network, full characterization of diffusion dynamics, in many cases, may not be necessary. This objective is achieved by creating a diffusion kernel that well-approximates standard diffusion models such as the susceptible-infected diffusion model, but lends itself to inversion, by design, via likelihood maximization or error minimization. We apply NI for both single-source and multi-source diffusion, for both single-snapshot and multi-snapshot observations, and for both homogeneous and heterogeneous diffusion setups. We prove the mean-field optimality of NI for different scenarios, and demonstrate its effectiveness over several synthetic networks. Moreover, we apply NI to a real-data application, identifying news sources in the Digg social network, and demonstrate the effectiveness of NI compared to existing methods.

**KEYWORDS:** Network Infusion, susceptible-infected, EGMP

## INTRODUCTION

Information from a single node (entity) can reach other nodes (entities) by propagation over network connections. For instance, a viral infection (either computer or biological) can propagate to different nodes in a network and become an epidemic, while rumors can spread in a social network through social interactions. Even a financial failure of an institution can have cascading effects on other financial entities and may lead to a financial crisis. As a final example, in some human diseases, abnormal activities of few encoding genes for example, transcription factors, can cause their target genes and therefore some essential biological processes to fail to operate normally in the cell. In order to gain insight into these processes, mathematical models have been developed, primarily focusing on application to the study of virus propagation in networks.

A well-established continuous-time diffusion model for viral epidemics is known as the susceptible-infected (SI) model, where infected nodes spread the virus to their neighbors probabilistically. For that diffusion model, explore the

relationship between network structure, infection rate, and the size of epidemics, while considering learning SI model parameters. Other diffusion methods use random walks to model information spread and label propagation in networks. These references study the forward problem of signal diffusion. Source inference is the inverse problem. It aims to infer source nodes in a network by merely knowing the network structure and observing the information spread at single or multiple snapshots (Figure 1). Even within the context of the well-studied diffusion kernels, source inference is a difficult problem in great part owing to the presence of path multiplicity in the network.

Network Infusion (NI) aims to identify source node(s) by reversing information propagation in the network. NI is based on a path-based network diffusion process that closely approximates the observed diffusion pattern, while leading to a tractable source inference method for large complex networks. The displayed network and infection pattern are parts of the Digg social news network. Recently, the inverse

**How to cite this paper:** K. Malarvizhi | Prof. P. Parthasarathy | Dr. S. Shankar "Data Passing with High Security through Non Infected Nodes in Networks" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-3 | Issue-6, October 2019, pp.1274-1278, URL: <https://www.ijtsrd.com/papers/ijtsrd29387.pdf>



IJTSRD29387

Copyright © 2019 by author(s) and International Journal of Trend in Scientific Research and Development Journal. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0) (<http://creativecommons.org/licenses/by/4.0>)



problem of a diffusion process in a network under a discrete time memory less diffusion model, and when time steps are known, has been considered, while the problem of identifying seed nodes (effectors) of a partially activated network in the steady state of an Independent-Cascade model is investigated in. Moreover reference has considered the source inference problem using incomplete diffusion traces by maximizing the likelihood of the trace under the learned model. The problem setup and diffusion models considered in those works are different than the continuous-time diffusion setup considered in the present paper. The work in uses the Minimum Description Length principle to identify source nodes using a heuristic cost function which combines the model cost and the data cost. Moreover, for the case of having a single source node in the network, some methods infer the source node based on distance centrality, or degree centrality measures of the infected sub graph. These methods can be efficiently applied to large networks, but, amongst other drawbacks, their performance lacks provable guarantees in general. For tree structures under a homogeneous SI diffusion model, computes a maximum likelihood solution for the source inference problem and provides provable guarantees for its performance. Over tree structures, their solution is equivalent to the distance centrality of the infected subgraph. The problem of inferring multiple sources in the network has an additional combinatorial complexity compared to the single-source case.

Reference has considered this problem under an Independent Cascade (IC) model and introduced a polynomial time algorithm based on dynamic-programming for some special cases. Source inference in real-world networks is made more challenging as the true propagation dynamics are typically unknown, and when they have been directly measured, they rarely conform to the assumptions of well-studied canonical diffusion models such as SI, partially owing to heterogeneous information diffusion over network edges, latent sources of information, noisy network connections, non-memory less transport and so forth. As an example, we consider news spread over the Digg social news network for more than 3,500 news stories. We find that in approximately 65% of cases, nodes who have received the news at a time  $t$ , did not have any neighbors who had already received the news by that time violating the most basic conditional independence assumption of the SI model. Furthermore, the empirical distribution of remaining news propagation times over edges of the Digg social news network cannot be approximated closely by a single distribution of a homogenous SI diffusion model, even by fitting a general Weibull distribution to the observed data.

Owing to high computational complexity of solving the source inference problem under the well-studied SI diffusion models and considering the fact that those kernels are unlikely to match precisely a real-world diffusion, our key idea to solve the inverse problem is to identify a diffusion process that closely approximates the observed diffusion pattern, but also leads to a tractable source inference method by design. Thus, we develop a diffusion kernel that is distinct from the standard SI diffusion models, but its order of diffusion well approximates many of them in various setups. We shall show that this kernel leads to an efficient source inference method that can be computed efficiently for large complex networks and shall provide theoretical

performance guarantees under general conditions. The key original observation, from both a theoretical and practical perspective, is that in order to solve the inverse problem one does not need to know the full dynamics of the diffusion, instead to solve the inversion one can do so from statistics that are consistent across many diffusion models.

## LITERATURE SURVEY

### 1. Statistical Inference of Computer Virus Propagation Using Non-Homogeneous Poisson Processes by H. Okamura, K. Tateishi, and T. Dohi,

This paper presents measurable surmising of PC infection spread utilizing non-homogeneous Poisson forms (NHPPs). Under some numerical suppositions, the quantity of tainted hosts can be displayed by a NHPP. Specifically, this paper applies a structure of blended sort NHPPs to the measurable deduction of intermittent infection spread. The blended sort NHPP is characterized by a superposition of NHPPs. In numerical analyses, we look at a decency of-fit measure of NHPPs on fitting to genuine infection disease information, and talk about the adequacy of the model-based expectation approach for PC infection spread. we have built up the factual models to portray the PC infection proliferation dependent on NHPPs. Specifically, when we apply the calculated and extraordinary esteem appropriations to the disease time dissemination, the subsequent mean conduct of NHPP models are actually the same as the outstanding strategic and Gompertz bends. Along these lines the structure of NHPP models basically contains the traditional relapse investigation. Additionally, we have presented the blended kind NHPP models to speak to the proliferation of PC infection. Since the blended kind NHPP models can express occasional disease wonder, it is better than the typical non-blended NHPPs with unimodal contamination time conveyance, as far as decency of-fit. For the factual examination, we have proposed the EM calculation with the goal that we could undoubtedly gauge show parameters for infection disease information. In numerical analyses, we have performed KS trial of the NHPP models for 116 sorts of infection information. Thus, all infection disease would be demonstrated by NHPPs. Additionally, we analyzed the expectation capacities for the proposed blended sort NHPP models, contrasted with the ordinary relapse models with the strategic and Gompertz bends. And we have explored that the blended sort NHPP models were fit for fitting to any sort of disease information in numerical examinations, the forecast capacity is inadequate to assess the future infection contamination notwithstanding when we utilize the blended kind NHPP.

### 2. Spotting Culprits in Epidemics: How many and Which ones? by B. A. Prakash, J. Vreeken, and C. Faloutsos,

Given a depiction of a substantial chart, in which a contamination has been spreading for quite a while, would we be able to recognize those hubs from which the disease begun to spread? At the ending of the day, can we dependably tell who the offenders are? Here, we answer this inquiry certifiably, and give a proficient technique called NETSLEUTH for the outstanding Susceptible-Infected infection proliferation display. Basically, we are after that arrangement of seed hubs that best clarify the given depiction. We propose to utilize the least Description Length guideline to distinguish the best arrangement of seed hubs and infection engendering swell, as the one by which we can

most compactly portray the contaminated diagram. We give an exceptionally productive calculation to distinguish likely arrangements of seed hubs given a preview. At that point, given these seed hubs, we demonstrate we can streamline the infection spread swell principally by augmenting probability. With every one of the three consolidated, NETSLEUTH can consequently recognize the right number of seed hubs, and additionally which hubs are the guilty parties. Experimentation on our strategy demonstrates high exactness in the discovery of seed hubs, notwithstanding the right programmed recognizable proof of their number. In addition, we indicate NETSLEUTH scales straightly in the quantity of hubs of the chart. we examined discovering guilty parties, the testing issue of distinguishing the hubs from which a contamination in a diagram began to spread. We proposed to utilize the Minimum Description Length standard for distinguishing that arrangement of seed hubs from which the given depiction can be portrayed generally concisely. We presented NETSLEUTH (in view of a novel 'submatrix-laplacian' strategy), a very proficient calculation for both distinguishing the arrangement of seed hubs that best portrays the given circumstance, and consequently choosing the best number of seed hubs—rather than the best in class.

### 3. The Effect of Network Topology on the Spread of Epidemics by A. Ganesh, L. Massoulié, and D. Towsley

Numerous framework ponders are all around shown as spreads of pandemics through a framework. Perceptible points of reference consolidate the spread of worms and email diseases, and, even more all things considered, weaknesses. Various sorts of information dispersal can similarly be shown as spreads of pandemics. In this paper we address the subject of what makes a pandemic either slight or serious. More precisely, we perceive topological properties of the chart that choose the creativity of sicknesses. In particular, we exhibit that if the extent of fix to pollution rates is greater than the unearthly scope of the graph, by then the mean pandemic lifetime is of demand  $\log n$ , where  $n$  is the amount of centers. Then again, if this extent is tinier than a theory of the isoperimetric predictable of the graph, by then the mean pandemic lifetime is of demand, for a positive relentless  $a$ . We apply these results to a couple of framework topologies including the hypercube, which is an agent accessibility graph for a spread hash table, the whole outline, which has a basic system graph for BGP, and the power law graph, of which the AS-level Inkmet diagram is a prime model.

We likewise ponder the star topology and the Erdős-Rényi chart as their plague spreading practices decide the spreading conduct of intensity law diagrams. We have displayed a primer examination of how topology influences the spread of a pestilence, roused by systems administration wonders, for example, worms and infections, falling disappointments, and dispersal of data. We have created adequate conditions under which pestilences either cease to exist rapidly (logarithmically in the span of the system) or gradually (exponentially in the measure of the system).

### 4. Explosive Percolation in Random Networks by D. Achlioptas, R. M. D'Souza, and J. Spencer

Systems in which the development of associations is represented by an arbitrary procedure regularly experience

a permeation change, wherein around a basic point, the expansion of few associations makes a sizable portion of the system all of a sudden wind up connected together. Ordinarily such advances are nonstop, with the goal that the level of the system connected together will in general zero appropriate over the progress point. Regardless of whether permeation advances could be spasmodic has been an open inquiry. Here, we demonstrate that fusing a restricted measure of decision in the great Erdős-Rényi arrangement development show causes its permeation change to end up spasmodic. Other than water transforming into ice or steam, other prototypical stage advances are the unconstrained development of charge and superconductivity in metals, the pestilence spread of malady, and the sensational change in availability of systems and cross sections known as permeation. Maybe the most key normal for a stage progress is its request, i.e., regardless of whether the plainly visible amount it influences changes persistently or spasmodically at the change. Persistent (smooth) changes are called second-order and incorporate numerous charge marvels, while spasmodic (unexpected) advances are called first-order, a commonplace model being the intermittent drop in entropy when fluid water transforms into strong ice at  $0^\circ\text{C}$ .

### 5. A tutorial introduction to Bayesian inference for stochastic epidemic models using Markov chain Monte Carlo methods by P. D. O'Neill,

Later Bayesian techniques for the investigation of irresistible ailment flare-up information utilizing stochastic plague models are surveyed. These strategies depend on Markov chain Monte Carlo techniques. Both fleeting and nontemporal information are considered. The strategies are outlined with various precedents highlighting diverse models and datasets. Contingent upon the application being referred to, models may join dormant periods, variable infectivity, diminished defenselessness following recuperation, and so forth. Also, parts of populace heterogeneity, for example, age structure, shifting helplessness, differential blending rates between gatherings of people, and so forth can be incorporated as suitable. Similarly as with any factual displaying, there is a harmony between models that are excessively convoluted for the information, making it impossible to completely illuminate, and those which are excessively shortsighted, making it impossible to be viewed truly as a reason for producing helpful induction. By and by it isn't constantly clear to accomplish this parity by means of a formal method; issues of model sufficiency and integrity of fit are not particularly all around created in the writing. Circumstances in which the information basically comprise of rehashed free perceptions.

### 6. Mixing patterns in networks by M. E. Newman

We contemplate assortative blending in systems, the propensity for vertices in systems to be associated with different vertices that resemble ~or not at all like! them here and there. We consider blending as per discrete qualities, for example, dialect or race in informal organizations and scalar attributes, for example, age. As an uncommon case of the last we consider blending as per vertex degree, i.e., as per the quantity of associations vertices need to different vertices: do gregarious individuals will in general connect with different gregarious individuals? We propose various proportions of assortative blending suitable to the different blending types, and apply them to an assortment of



certifiable systems, demonstrating that assortative blending is an unavoidable wonder found in numerous systems. We additionally propose a few models of assortatively blended systems, both explanatory ones dependent on producing capacity strategies, and numerical ones dependent on Monte Carlo diagram age techniques. We utilize these models to test the properties of systems as their dimension of assortativity is shifted. In the specific instance of blending by degree, we find solid variety with assortativity in the availability of the system and in the flexibility of the system to the evacuation of vertices. Assortative blending can profoundly affect the basic properties of a system. For instance, assortative blending of a system by a discrete trademark will in general split the system up into independent networks. In the event that individuals want to be companions with other people who talk their very own dialect, for instance, at that point one may expect nations with in excess of one dialect to isolate into networks by dialect. Assortative blending by age could cause stratification of social orders along age lines.

## REFERENCES

- [1] R. Pastor-Satorras and A. Vespignani, "Epidemic spreading in scalefree networks," *Physical review letters*, vol. 86, no. 14, p. 3200, 2001.
- [2] D. Acemoglu, A. Ozdaglar, and A. ParandehGheibi, "Spread of (mis) information in social networks," *Games and Economic Behavior*, vol. 70, no. 2, pp. 194–227, 2010.
- [3] D. Hirshleifer and S. Hong Teoh, "Herd behaviour and cascading in capital markets: A review and synthesis," *European Financial Management*, vol. 9, no. 1, pp. 25–66, 2003.
- [4] M. T. Maurano, R. Humbert, E. Rynes, R. E. Thurman, E. Haugen, H. Wang, A. P. Reynolds, R. Sandstrom, H. Qu, J. Brody et al., "Systematic localization of common disease-associated variation in regulatory dna," *Science*, vol. 337, no. 6099, pp. 1190–1195, 2012.
- [5] A. B. Glinskii, J. Ma, S. Ma, D. Grant, C.-U. Lim, S. Sell, and G. V. Glinsky, "Identification of intergenic trans-regulatory rnas containing a disease-linked snp sequence and targeting cell cycle progression/differentiation pathways in multiple common human disorders," *Cell Cycle*, vol. 8, no. 23, pp. 3925–3942, 2009.
- [6] T. W. Valente, "Network models of the diffusion of innovations," *Computational and Mathematical Organization Theory*, vol. 2, no. 2, pp. 163–164, 1996.
- [7] D. Achlioptas, R. M. D'Souza, and J. Spencer, "Explosive percolation in random networks," *Science*, vol. 323, no. 5920, pp. 1453–1455, 2009.
- [8] N. T. Bailey et al., *The mathematical theory of infectious diseases and its applications*. Charles Griffin., 1975.
- [9] M. E. Newman, "Spread of epidemic disease on networks," *Physical review E*, vol. 66, no. 1, p. 016128, 2002.
- [10] C. Moore and M. E. Newman, "Epidemics and percolation in small world networks," *Physical Review E*, vol. 61, no. 5, p. 5678, 2000.
- [11] A. Ganesh, L. Massoulié, and D. Towsley, "The effect of network topology on the spread of epidemics," in *24th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM)*, vol. 2, 2005, pp. 1455–1466.
- [12] N. Demiris and P. D. O'Neill, "Bayesian inference for epidemics with two levels of mixing," *Scandinavian journal of statistics*, vol. 32, no. 2, pp. 265–280, 2005.
- [13] P. D. O'Neill, "A tutorial introduction to bayesian inference for stochastic epidemic models using markov chain monte carlo methods," *Mathematical Biosciences*, vol. 180, no. 1, pp. 103–114, 2002.
- [14] H. Okamura, K. Tateishi, and T. Dohi, "Statistical inference of computer virus propagation using non-homogeneous poisson processes," in *IEEE International Symposium on Software Reliability*, 2007, pp. 149–158.
- [15] S. Mostafavi, A. Goldenberg, and Q. Morris, "Labeling nodes using three degrees of propagation," *PloS one*, vol. 7, no. 12, p. e51947, 2012.
- [16] Y. Bengio, O. Delalleau, and N. Le Roux, "Label propagation and quadratic criterion," *Semi-supervised learning*, pp. 193–216, 2006.
- [17] M. E. Newman, "Mixing patterns in networks," *Physical Review E*, vol. 67, no. 2, p. 026126, 2003.
- [18] E. Sefer and C. Kingsford, "Diffusion archaeology for diffusion progression history reconstruction," *IEEE International Conference on Data Mining (ICDM)*, pp. 530–539, 2014.
- [19] M. Farajtabar, M. Gomez-Rodriguez, N. Du, M. Zamani, H. Zha, and L. Song, "Back to the past: Source identification in diffusion networks from partially observed cascades," *arXiv preprint arXiv:1501.06582*, 2015.
- [20] B. A. Prakash, J. Vreeken, and C. Faloutsos, "Spotting culprits in epidemics: How many and which ones?" in *International Conference on Data Mining (ICDM)*, 2012, pp. 11–20.
- [21] M. Newman, *Networks: an introduction*. Oxford University Press, 2010.
- [22] D. B. West, *Introduction to graph theory*. Prentice hall Upper Saddle River, 2001, vol. 2.
- [23] D. Shah and T. Zaman, "Rumors in a network: Who's the culprit?" *IEEE Transactions on Information Theory*, vol. 57, no. 8, pp. 5163–5181, 2011.
- [24] T. Lappas, E. Terzi, D. Gunopulos, and H. Mannila, "Finding effectors in social networks," in *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2010, pp. 1059–1068.
- [25] N. Du, L. Song, H. Woo, and H. Zha, "Uncover topic-sensitive information diffusion networks," in *Proceedings of the sixteenth international conference on artificial intelligence and statistics*, 2013, pp. 229–237.
- [26] S. Goel, D. J. Watts, and D. G. Goldstein, "The structure of online diffusion networks," in *Proceedings of the*

- 13th ACM conference on electronic commerce. ACM, 2012, pp. 623–638.
- [27] K. Lerman and R. Ghosh, "Information contagion: An empirical study of the spread of news on digg and twitter social networks." ICWSM, vol. 10, pp. 90–97, 2010.
- [28] J. Hershberger, M. Maxel, and S. Suri, "Finding the k shortest simple paths: A new algorithm and its implementation," ACM Transactions on Algorithms (TALG), vol. 3, no. 4, p. 45, 2007.
- [29] C. De Bacco, S. Franz, D. Saad, and C. H. Yeung, "Shortest nodedisjoint paths on random graphs," arXiv preprint: 1401.8096, 2014.
- [30] P. Erdős and A. Rényi, "On the strength of connectedness of a random graph," Acta Mathematica Hungarica, vol. 12, no. 1, pp. 261–267, 1961.
- [31] D. P. Bertsekas, Introduction to Probability: Dimitri P. Bertsekas and John N. Tsitsiklis. Athena Scientific, 2002.
- [32] R. Kohavi et al., "A study of cross-validation and bootstrap for accuracy estimation and model selection," in IJCAI, vol. 14, no. 2, 1995, pp. 1137–1145.
- [33] M. G. Rodriguez, J. Leskovec, D. Balduzzi, and B. Schölkopf, "Uncovering the structure and temporal dynamics of information propagation," Network Science, vol. 2, no. 01, pp. 26–65, 2014.
- [34] E. Sefer and C. Kingsford, "Convex risk minimization to infer networks from probabilistic diffusion data at multiple scales," in IEEE International Conference on Data Engineering (ICDE), 2015.

